University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Fall 9-16-2019

Bibliometric Survey on Incremental Clustering Algorithms

Archana Chaudhari
Research Scholar, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India., archana.chaudhari@sitpune.edu.in

Rahul Raghvendra Joshi
Research Scholar and Assistant Professor, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India., rahulj@sitpune.edu.in

Preeti Mulay
Ph.D. Guide and Associate Professor, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India., preeti.mulay@sitpune.edu.in

Ketan Kotecha
Director, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India., director@sitpune.edu.in

Parag Kulkarni
Innovation Strategist, CEO, and Chief Scientist, iknowlation Research Labs Pvt Ltd Pune, India, paragindia@gmail.com

Chaudhari, Archana; Joshi, Rahul Raghvendra; Mulay, Preeti; Kotecha, Ketan; and Kulkarni, Parag, "Bibliometric Survey on Incremental Clustering Algorithms" (2019). *Library Philosophy and Practice (e-journal)*. 2762.
https://digitalcommons.unl.edu/libphilprac/2762

Bibliometric Survey on Incremental Clustering Algorithms

Archana Chaudhari[1], Rahul Raghvendra Joshi[2], Dr. Preeti Mulay[3], Dr. Ketan Kotecha[4], Dr. Parag Kulkarni[5]

[1]*Research Scholar, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.*

Email: archana.chaudhari@sitpune.edu.in

[2]*Research Scholar and Assistant Professor, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.*

Email: rahulj@sitpune.edu.in

[3]*Ph.D. Guide and Associate Professor, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.*

Email: preeti.mulay@sitpune.edu.in

[4]*Director, Symbiosis Institute of Technology (SIT) affiliated to Symbiosis International (Deemed University), Pune, India.*

Email: director@sitpune.edu.in

[5]*Innovation Strategist, CEO, and Chief Scientist, iknowlation Research Labs Pvt Ltd Pune, India.*

Email: paragindia@gmail.com

ABSTRACT

For clustering accuracy, on influx of data, the parameter-free incremental clustering research is essential. The sole purpose of this bibliometric analysis is to understand the reach and utility of incremental clustering algorithms. This paper shows incremental clustering for time series dataset was first explored in 2000 and continued thereafter till date. This Bibliometric analysis is done using Scopus, Google Scholar, Research Gate, and the tools like Gephi, Table2Net, and GPS Visualizer etc. The survey revealed that maximum publications of incremental clustering algorithms are from conference and journals, affiliated to Computer Science, Chinese lead publications followed by India then United States. Convergence optimality is another prominent keyword and less attentiveness towards correlation has observed. For betweenness and friendly measures keywords, after physics and astronomy; engineering is the contributing subject area,

minimal contribution of review papers are observed in this art-search. The effectual incremental learning is feasible via parameter-free incremental clustering algorithm, applicable to all domains and hence this study.

Keywords: Bibliometric Analysis, Incremental Clustering, CFBA, Machine Learning

1. INTRODUCTION

Data is continuously generated through various important applications viz. monitoring of financial transactions, intelligent energy network flow, satellite imagery, and information through web processing. Data mining techniques and results evolve with this kind of newly generated data. It necessitates an incremental update of new clusters on previous records rather than to re-cluster complete data from the beginning. The problem of clustering associated with growing data can be resolved using incremental clustering. This approach creates incremental learning to result in knowledge augmentation (Mulay & Kulkarni, 2013). The augmented knowledge can be useful to develop a novel strategy for clustering. The augmented knowledge can be derived using closeness among data points. Closeness can be determined using two ways viz. distance based and pattern based (Prachi M Joshi & Kulkarni, 2011). Distance-based closeness between patterns may not be suitable for all applications. It would lead to the curse of dimensionality. Pattern-based closeness factor compares the pattern of occurrences. It represents a thematic relationship and coherence between one or more objects. Closeness Factor (CF) may be activity specific or decision specific. The working principle of Algorithm based on Closeness Factor (CFBA) (Kulkarni & Mulay, 2013) is closer the data points; higher is the probability that they belong to the same chunk called a cluster. The CF value quantifies disparity in the form of data series. CF value equal to zero signifies that data series exactly match with each other even though volumes might be different (Kulkarni, Dwivedi, & Haribhakta, 2015). CF-based algorithm acts as an enhancer for knowledge augmentation process (Gaikwad, Joshi, &Mulay, 2016). CFBA creates a new cluster if new information does not match with the already formed clusters (Swamy& Kulkarni, 2006). Learning considers the behavioral pattern of data — all-time active learning required for getting the knowledge base evolved (Archana Chaudhari & Mulay, 2018) of clustering algorithms. CFBA modifies knowledge and maintains patterns for reuse. This modified knowledge and patterns saves clustering time and helps in decision-making. While the learning takes place, the cluster quality is also maintained. Thus, CFBA can play a vital role in a scenario where dynamic learning is manifested (Johnson & Singh, 2016). Table 1 shows advancements in single machine versions of CFBA.

Table 1: Advancements in the desktop versions of CFBA

Sr. No.	Evolved versions of CFBA	Year	Datasets used
1	Enhanced Closeness Factor Algorithm for Effectual Forecasting (ECFAEF)	2010	Organizational database
2	Probabilistic CFBA (P. Mulay & Kulkarni, 2013)	2013	Wine, Software Development Life Cycle
3	Modified Cluster Formation Algorithm (MCFA) (Suhas M. G, Rahul R. Joshi, & Mulay, 2015)	2015	Ice cream, Diabetes patients datasets
4	Incremental Clustering Naïve Bayes Closeness-factor Algorithm (ICNBCFA)	2016	Wine, Electricity, Software Project, Zenni Optics and Wine quality
5	Correlation-based Incremental Clustering Algorithm (CBICA) (Mulay &Shinde, 2017)	2017	Wine
6	Threshold Based Clustering Algorithm (TBCA) (Mulay, Joshi, et al., 2017)	2017	Diabetes Mellitus patients datasets
7	Deep Incremental Statistical Closeness Factor-Based Algorithm (DIS-CFBA) (Joshi & Mulay, 2018)	2018	Diabetes Mellitus patients datasets

The ECFAEF is the first version of CFBA. ECFAEF formulated probability, expected value, error and weight terminologies related to data series. Probabilistic CFBA compared with IK-Means and cobweb to gauge its performance. Probabilistic CFBA outperformed over these two incremental clustering algorithms. In MCFA Manhattan distance used for cluster formation. ICNBCFA does threshold computation using Naïve Bayes. CBICA is probability free approach. CBICA replaced probability-based calculation in CFBA by Pearson's correlation coefficient. TBCA used threshold, cluster average versus outlier average, cluster deviation versus outlier deviation for identification of impactful attributes. All these versions of CFBA follow parameter-free clustering approach.

The properties that make CFBA incremental are:

- Cluster-primary approach
- Iterative convergence
- Error based clustering
- Acquiring knowledge and augmentation
- Cluster-class assignment
- Ease of execution

CFBA, TBCA, and CBICA process numeric and mixed type of data sets. The raw input data formats that can be supplied as an input to these variants are:

- Time-series
- Boolean
- Spatial-temporal
- Alphanumeric

CFBA and its variants are order independent algorithms as patterns in obtained clusters are reordered to match with patterns of already stored clusters. Because of this property, these algorithms can effectively handle unstructured data sets also. The increase in the number of attributes will increase the volume of the data set. In such a case, it becomes difficult to find apt patterns. Hence, to find out patterns with related attributes, the data set needs to be cleansed in the pre-clustering phase. The cleansing can be done using Principal Component Analysis (PCA), Sparse Principal Component Analysis (SPCA) (H. Kulkarni, 2017), Independent Component Analysis (ICA), Feature Selection (FS), Feature Extraction (FE) and to name a few. CFBA and all other its variants follow similar kind of approach for the identification of correct patterns. In case of distance based pattern matching the following three examples will be possible through all objects have the same pattern

1. Distance wise they are close to each other
2. Bottom objects are nearer to each other
3. Objects are shifted from each other

Therefore, even though objects have a similar pattern in all of the above three cases, objects are also scaled. This research paper is intended to show an assemblage of objects having the same patterns into clusters. Hence, in this regard, CFBA and its subsequent advancements found to be superior for closeness based pattern matching. Calculation of CF fails if the weight of data series is negative. Weight reflects relative importance. The ionosphere dataset from the UCI repository is the example of negative weight for CF. The weight of data series in almost all types of data sets is positive. Closeness gives similarity information about metadata of the dataset. The advantageous point about the use of CFBA and its related advancements is that their working without prior information about clusters (A. Kulkarni, Tokekar, & Kulkarni, 2015). Another advantage is that they capture the behavioral pattern of data series by association and dissociation between data point's patterns. Similarity algorithms mainly capture unsupervised information (Mulay, Patel, & Gauchia, 2017). Pattern-based clustering is commonly used for gene expression or chromosome matching; it is uncommon in the area of data mining. Thus, if we look at positive and negative correlation or CF values for data series, then the similarity is achieved through learning a boundary that separates positive and negative entities. Here, positive entities correspond to most likely dimensions and exhibit relevance feedback.

On the other hand, negative entities correspond to minimum likely dimensions. But, the drawback of these approaches is that they consider attribute space globally and dependencies between features ignored. A correlation-based similarity measure can be useful to produce concept clusters useful for trend analysis and future assumptions prediction. The ranking is based on correlation and CF values of data series.

Semi-supervised mode of learning allows learning from both labeled and unlabeled data. It is other way filters relevant unlabelled data. Semi-supervised incremental learning builds new learning parameters at the same time retains useful knowledge build in the past (Sheshang D. Degadwala et al. 2017). CFBA and its related advancements perform selective incremental and semi-supervised learning. This approach updates old learning parameters and at the same time builds a new one. This is also useful for hypothesis checking for new learning and information retrieval in a multidisciplinary approach (Zhongyu, 2013). Therefore, knowledge refinement and knowledge building is the necessity for today and future too.

This is all about developments in incremental clustering through the CFBA. To get a broader scenario about incremental clustering research area, there is a need of bibliometric survey of incremental clustering algorithms. This paper presents a bibliometric study of incremental clustering algorithms through section 2 highlighting preliminary data collection about incremental clustering publications; section 3 giving thorough bibliometric analysis. Future scope and limitations of the undertaken research are in sections 4 and 5 respectively followed by concluding remarks in section 6 and reference at the end of this paper.

2. PRELIMINARY DATA COLLECTION

There are two categories to access the publication databases namely, open access and paid access (Sarmiento, & Nagi, 1999). One can access these publications through their university library portals or by registering separately on individual websites (Sajeev Kadama et al. 2016). Also, there are several popular methods to retrieve the data from required databases. Popular publication databases are Scopus, Clarivate, SCImago, Mendeley, ScienceDirect, DBLP, Google Scholar, and Research Gate etc. Scopus is the largest abstract and citation database of peer-reviewed research literature in the field of science, engineering, technology, medicine, social sciences, arts, and humanities. The paper considers the Scopus database with the help of significant keywords identified in section 2.1.

2.1 Significant keywords

The crucial keywords related to incremental clustering were segregated into two compartments viz., master and primary type. For this research, table 2 enumerates keywords list used as a search tactic.

Table 2: Planned search tactic for Keywords

Master-Keyword	"incremental clustering."
Primary-Keyword (AND)	“closeness factor “OR “threshold” OR “friendly measures” OR “correlation” OR “Nearness” OR “incremental learning” OR “NaïveBayes" OR “probability” OR “log likelihood” OR “total friends” OR “proximity” OR “medoids” OR “betweeness” OR “similarity”

2.2 Initial search results

Scopus database is the base of this research paper. Preliminary investigation through planned keywords search tactic generated in all 173 publications. This is then restricted to 160 English publications only (Table 3).

Table 3: Incremental clustering publishing languages trends

Sr. No.	Publishing Language	Publications count
1	English	160
2	Chinese	12
3	French	1
Total		173

Source: http://www.scopus.com (accessed on 19th July 2018)

All kinds of published and unpublished publications considered. The researchers in incremental clustering research area have publicized recent papers in conferences. 53.76% of conference proceedings papers were there and 42.77% of journals and articles were there (Table 4).

Table 4: The Publication types in incremental clustering

Publication Type	Number of publications	Percentage of 173
Conference Paper	93	53.76%
Journal and Article	74	42.77%
Conference Review	4	2.31%
Conference Proceedings	3	1.73%
Book Series	2	1.16%
Total	173	100%

Source: http://www.scopus.com (accessed on 19th July 2018)

2.3 Preliminary data highlights

The related documents retrieved as journal papers, conference papers, articles, reports, etc. for the span of eighteen years from 2001 to 2018. Trends for yearly publication are shown in table 5 and in figure 1 for incremental clustering research area.

Table 5: Trends for Yearly publishing in incremental clustering

Year	Publications count	Year	Publications count
2018	2	2009	9
2017	12	2008	10
2016	15	2007	5
2015	13	2006	8
2014	16	2005	9
2013	14	2004	2

2012	12	2003	3
2011	17	2002	5
2010	17	2001	4
Total		173	

Source: http://www.scopus.com (accessed on 19th July 2018)

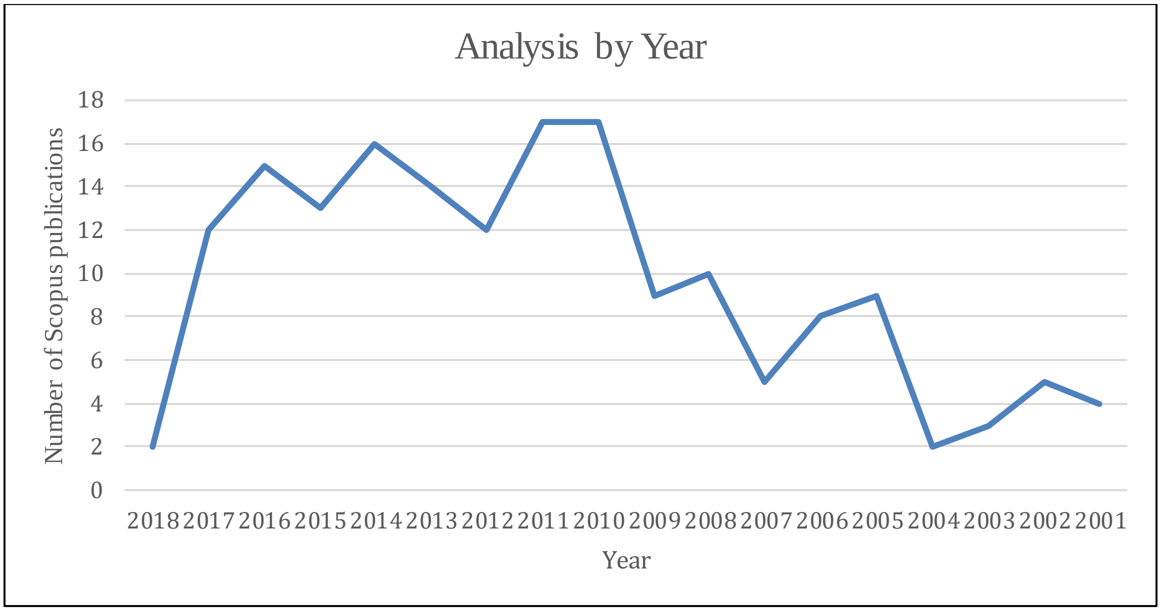


Figure 1: Yearly publishing trend in incremental clustering

2.4 Data Investigation

The thorough bibliometric study was conducted in section 3 to know the distinctiveness of literature, prominent researches and the researchers in the incremental clustering area through geographical attentiveness of the research, affiliation statistics, author contributions, journals where papers have been published and their statistics, along with an analysis of citation and collaborative studies.

3. BIBLIOMETRIC ANALYSIS

To perform a bibliometric analysis of incremental clustering concept, the following two ways are applied and they are:

- Analysis of geographic region, network, and citation etc.
- Statistics about the keyword, affiliation, author and journal, to name a few.

3.1 Geographical regional analysis

Figure 2 is drawn using gpsvisualiser.com showing geographical regional location attentiveness of published papers. The concentration of origin depends on the size of the circle. It is evident from the map that most of the researchers are from European countries.

Figure 2: Geographic locations of incremental clustering studies

Figure 3 gives the first ten countries having publications in the area of incremental clustering. It can be without a doubt that Chinese lead with nearly 35% followed by India with 17% and the United States (US) with 11% in publications.

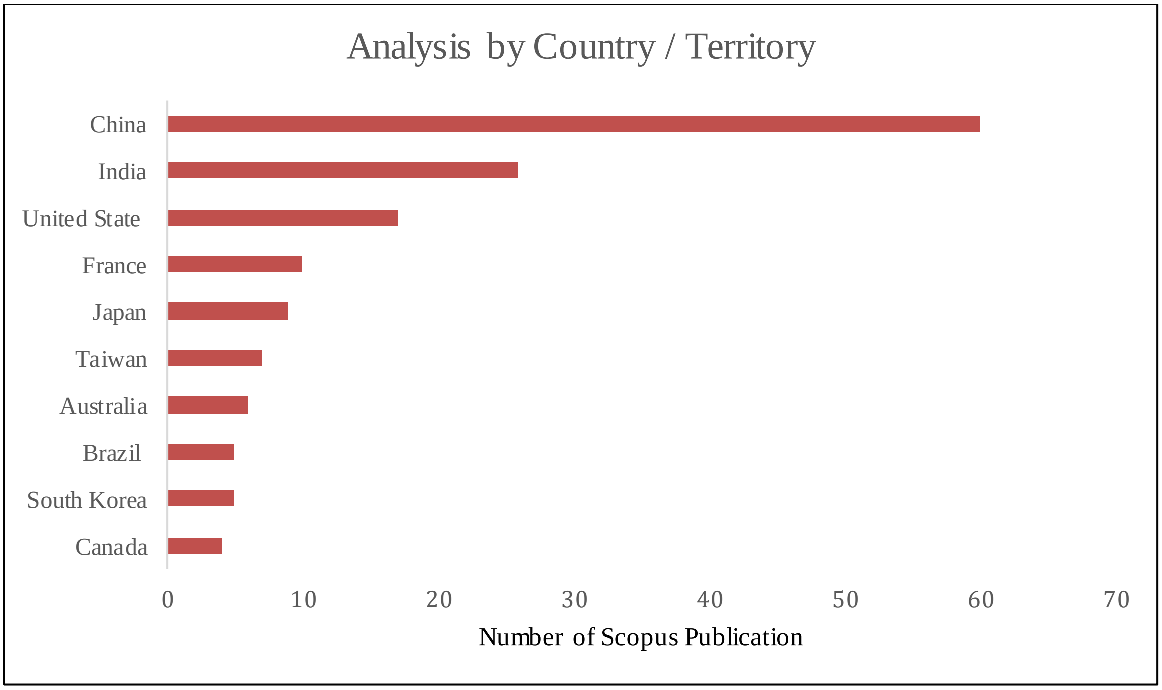


Figure 3: Top ten countries publishing papers on incremental clustering

Source: http://www.scopus.com (accessed on 19th July 2018)

3.2 Keywords statistics

Apt keywords indicate what the researchers want to search. The correct amalgamation of keywords helps to target significant research areas. First ten keywords list is there in Table 6 from considered publications in incremental clustering.

Table 6: First ten keywords for incremental clustering

Keywords	Number of Publications
Incremental Clustering	34
Clustering Algorithms	25
Cluster Analysis	9
Data Mining	8
Clustering	7
Artificial Intelligence	4
Database Systems	4
Learning System	4
Probability	3

Neural Network	2

Source: http://www.scopus.com (accessed on 19th July 2018)

3.3 Network Analysis

The relationship among different statistical parameters can be presented graphically using network analysis. "Gephi" is open source software. Gephi facilitates filtering, navigation, manipulation and clustering of network data. Different Authors, keywords used by them, citations received, their affiliations, and publication title, its year are shown using nodes and edges. In this paper for layout, Fruchterman Reingold and Yifan Hu Proportional were used with different manual adjustments. Networks with different parametric combinations for an incremental clustering algorithm for the data extracted from Scopus are shown in Figures 4–9. In figure 4 there are three essential data clusters for a cluster of author keywords and source title having 167 nodes and 169 edges.

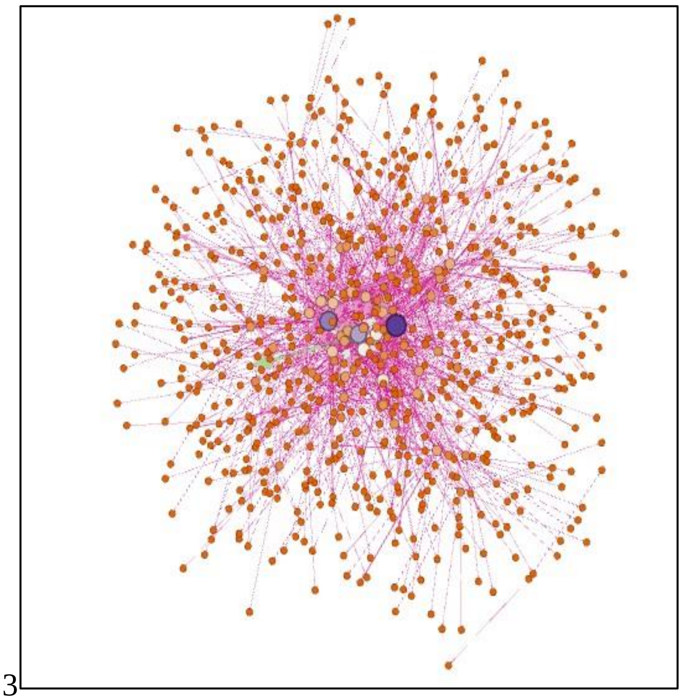
3


Figure 4: Cluster of author keywords and source title

A cluster of publication title, their publication year is shown in figure 5. The node size indicates utmost publications are mostly from 2010-2016. The network in figure 5 is having 201 nodes and 183 edges.

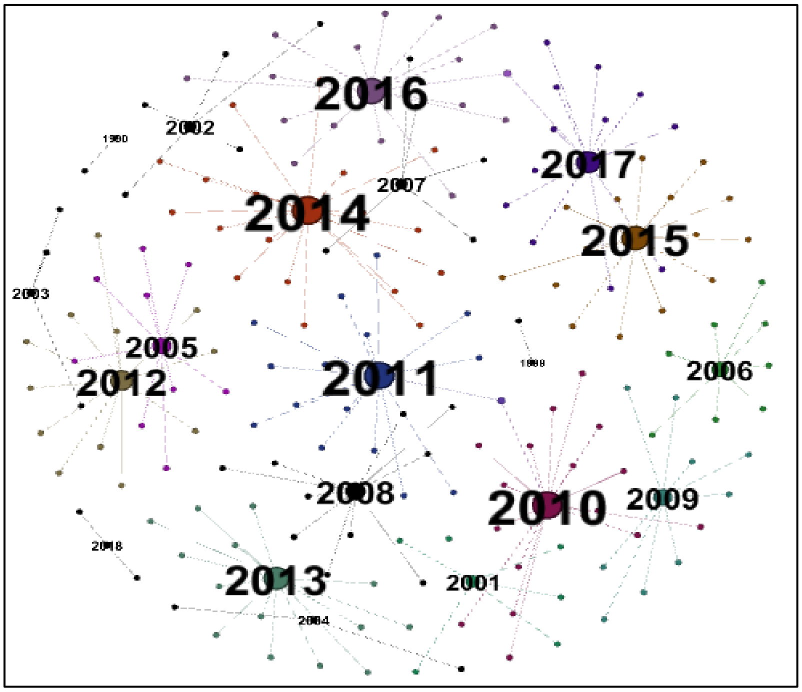


Figure 5: Cluster of Publication title and year of publication

Figure 6 shows a network of authors and author's keywords co-appearing in the same paper. Initial layout with 99 nodes and 62 edges. It has been observed that 'incremental clustering' and 'convergence optimality' are two major author keywords which are used extensively in this area. Also, it shows the application of incremental clustering algorithms in the area of frequent mining, document clustering, dynamic learning etc.

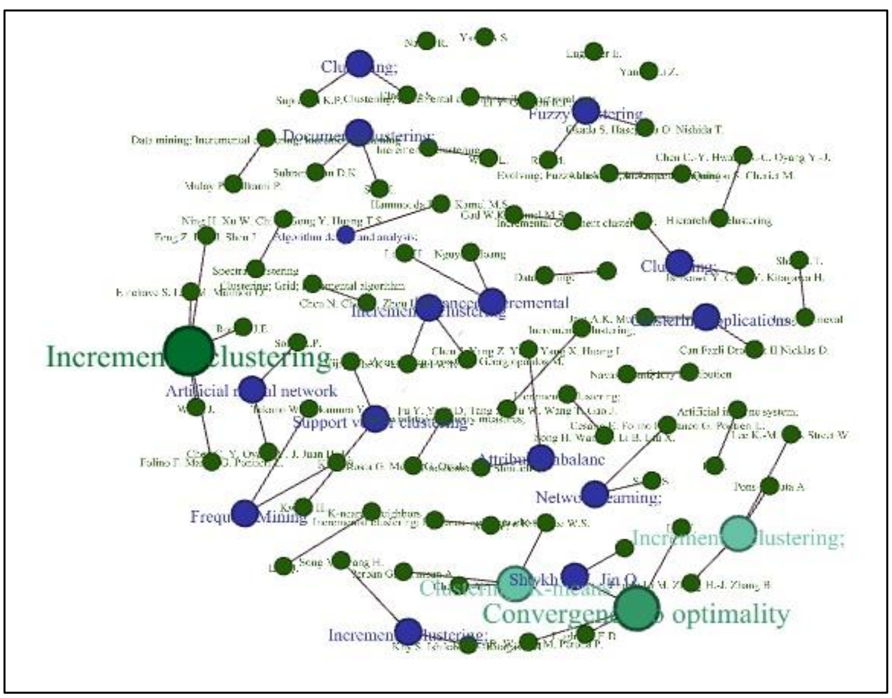


Figure 6: Cluster of authors, their keywords, co-appearance among similar papers

Figure 7 shows clusters in the data while clustering networks with author keywords and source title with 471 nodes and 351 edges.

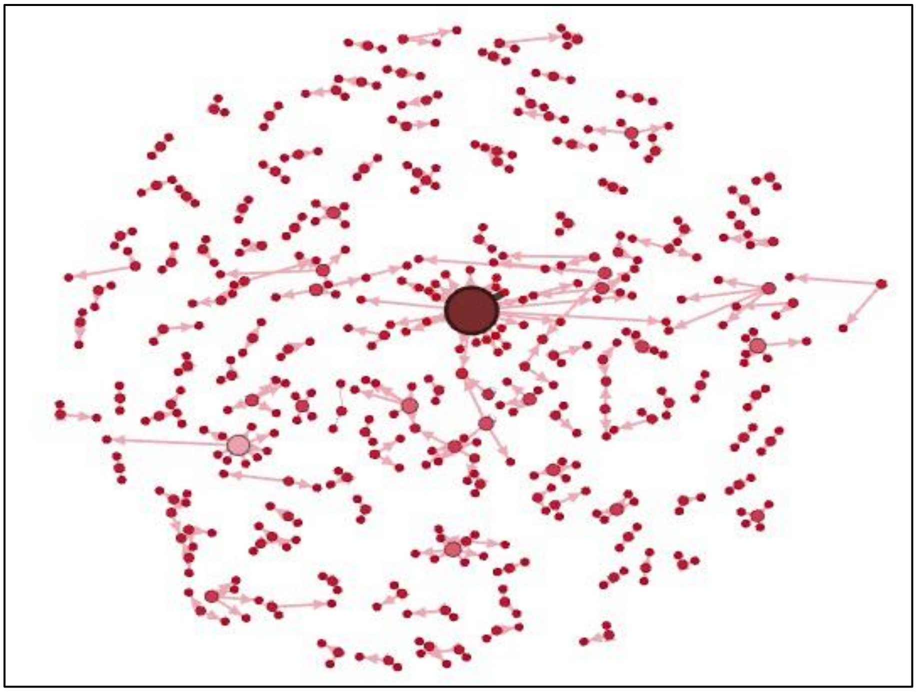


Figure 7: Cluster of authors and source titles, co-appearing among the same papers

Figure 8 shows that a cluster of affiliation, language, and type with 179 nodes and 340 edges. The size of the node indicates proportionate affiliation, language, and type. Maximum numbers of publications are published in 'Computer Science' affiliations

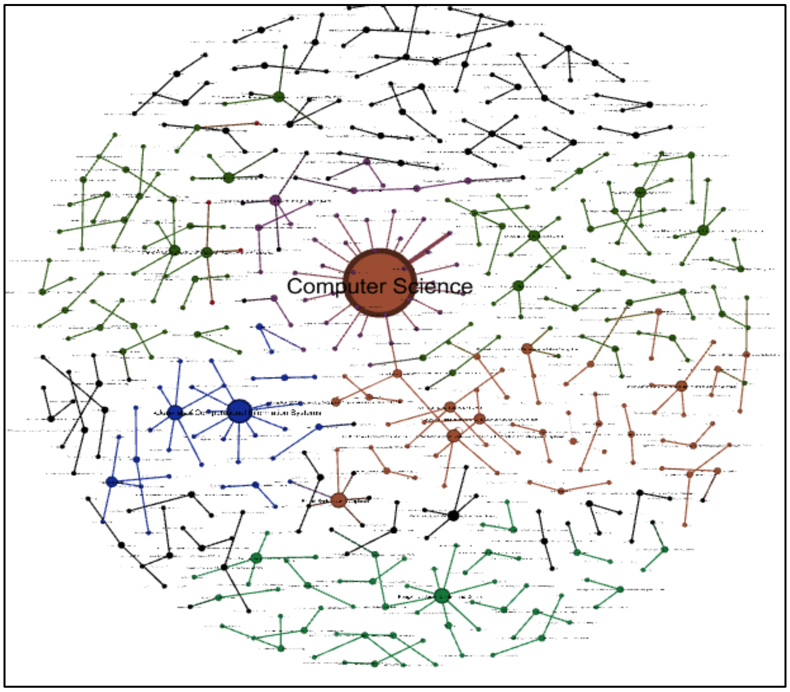


Figure 8: A cluster of affiliation, language, and type

Figure 9 shows a network of publication title and the number of citations received by publications published in it. The layout shows 32 nodes and 48 edges. The color of the nodes and labels indicates some citation linkage by publication.

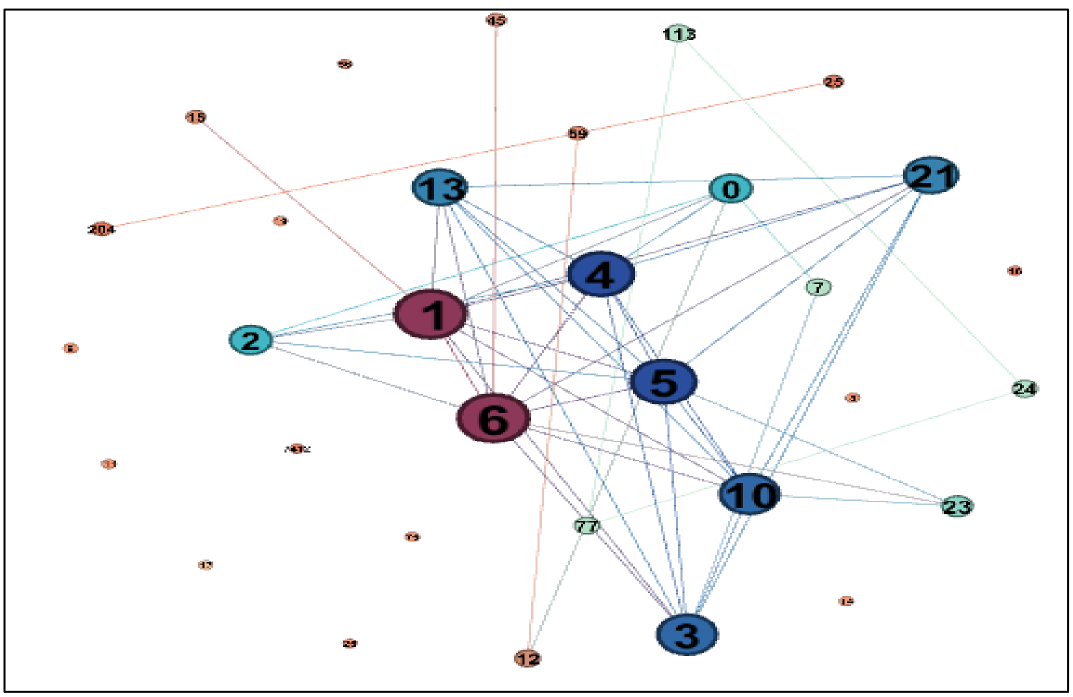


Figure 9: Publication title and number of citations cluster

3.4 Subject Areas

Figure 10 shows subject area wise compartmentalization for extracted incremental clustering publications. It is clear from this figure that maximum research is carried out in physics and astronomy followed by engineering and mathematics.It is also observed that less amount of research been carried out in the area of materials science (e.g. solar panel related research) and medicine (e.g. diabetes, cancer-related studies) which can act as a tool for human welfare.

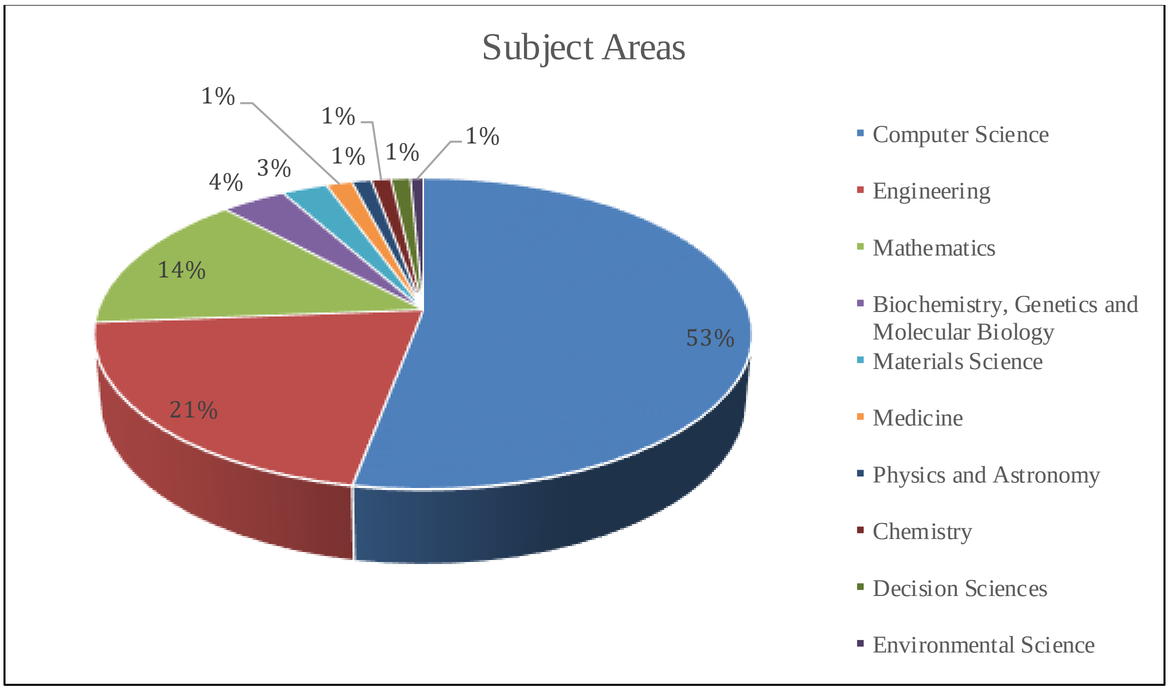


Figure 10: Subject area of extracted literature for incremental clustering

Source: http://www.scopus.com (accessed on 19th July 2018)

3.5 Affiliation statistics

Figure 11 indicates the top ten contributing universities/organizational affiliations. The incremental clustering has been an area of research concern among Chinese Academy of Sciences and Institute of Software Chinese Academy of Sciences. Interestingly six out of the top ten universities or institutions contributing to incremental clustering are from China.

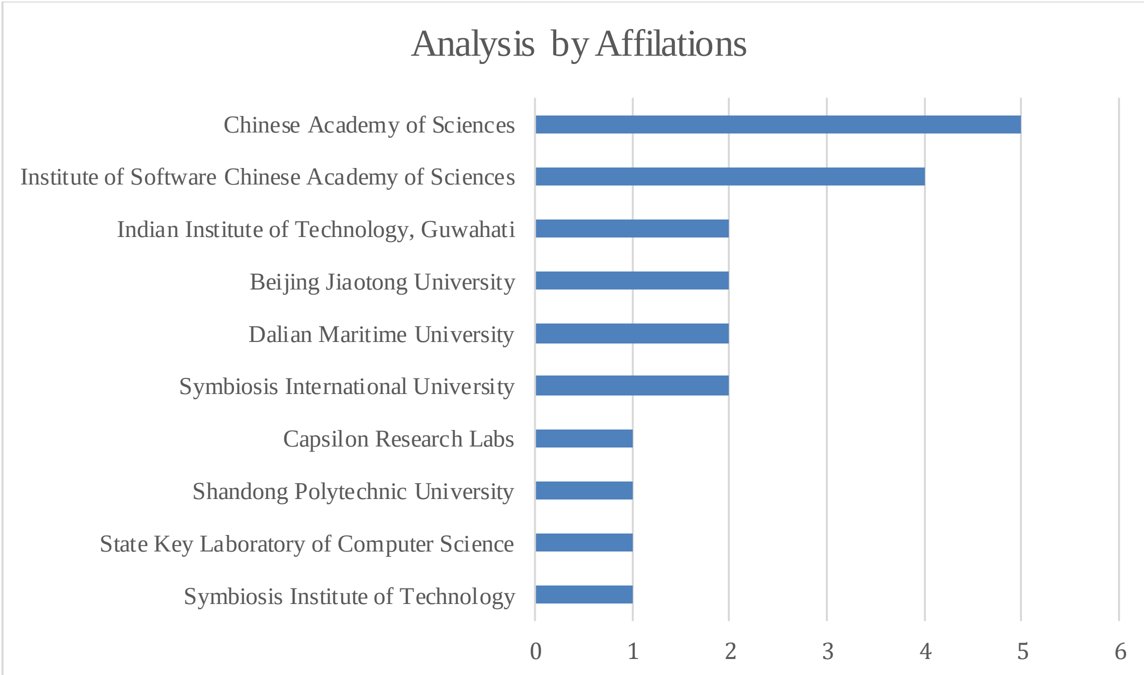


Figure 11: Affiliation statistics for incremental clustering

Source: http://www.scopus.com (accessed on 19th July 2018)

3.6 Author statistics

Figure 12 depicts the top ten authors contributing in the area of incremental clustering to understand the influence of a particular author.

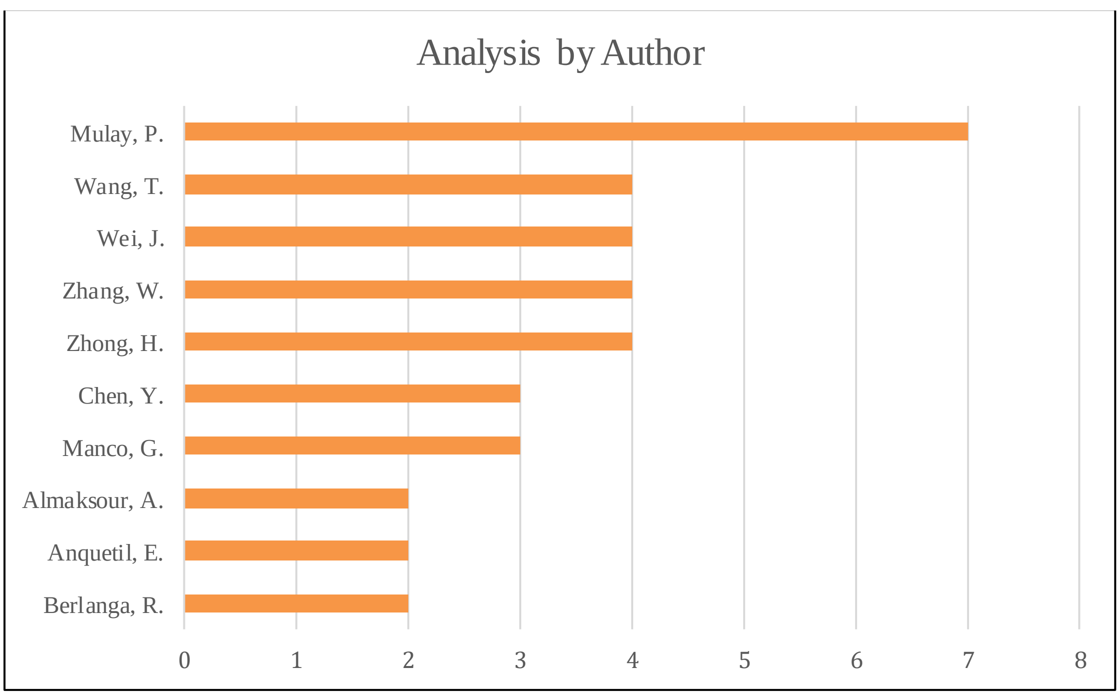


Figure 12: Key contributing authors

Source: http://www.scopus.com (accessed on 19th July 2018)

3.7 Journal statistics

Figure 13 covers the publication source types in the area of incremental clustering. The extracted information demonstrates that 95% of publications are either from book series or articles. Also, it is observed that only 2% of contribution in the form of review publications. Therefore, this bibliometric review on incremental clustering algorithms is written.

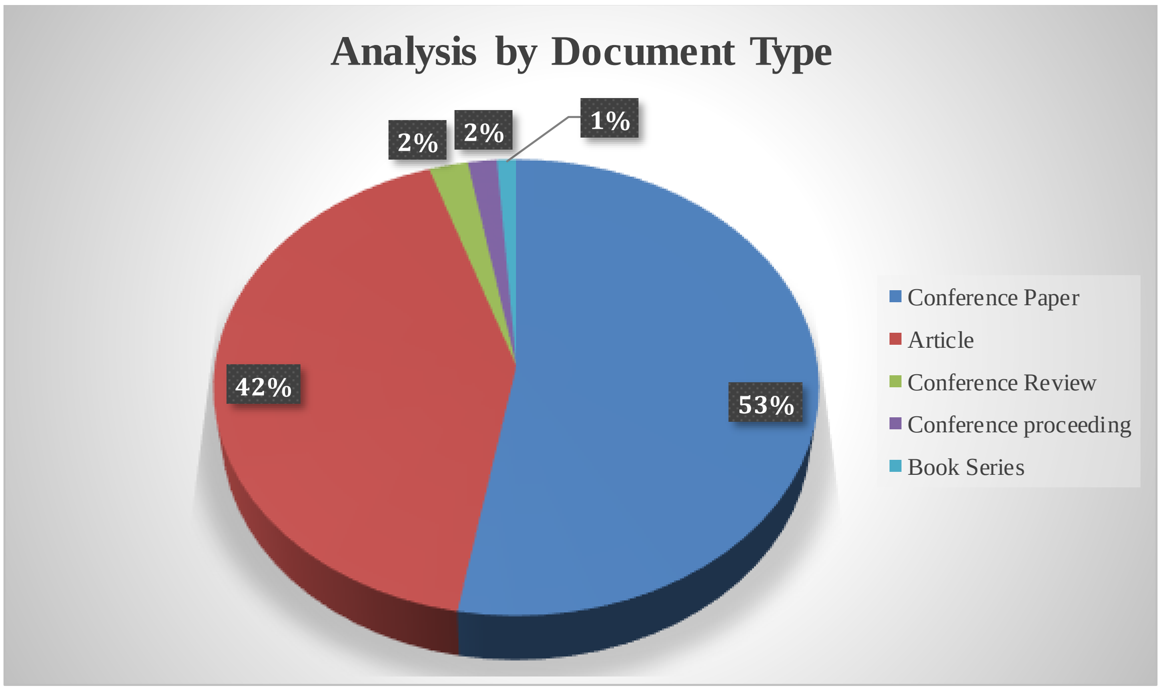


Figure 13: Types of sources for publications in incremental clustering

Source: http://www.scopus.com (accessed on 19th July 2018)

3.8 Citation Analysis

Table 7 shows yearly citations obtained through publications extracted in the area of incremental clustering. The total citation count of 173 publications is 6690 to date. The list of the first ten papers along citations received to them till the date of the data extracted for this research is in Table 8.

Table 7: Analysis of citations for publications in incremental clustering

Year	<2010	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total

No. of citations	640	627	655	672	750	735	724	744	743	370	6690

Source: http://www.scopus.com (accessed on 19th July 2018)

Table 8: A citation analysis of top ten publications in incremental clustering

Publication Title	Yearly citations received by the publication										
	<2010	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Extensions of vector quantization for incremental clustering	15	14	20	13	13	10	12	7	5	4	113
Incremental spectral clustering by efficiently updating the eigen system	-	1	5	10	11	17	6	17	6	3	77
Mode-seeking by medoidshifts	4	7	13	12	11	6	3	8	6	5	75
Ellipsoid ART and ARTMAP for incremental clustering and classification	38	5	3	2	0	3	3	1	2	2	59
Incremental fuzzy clustering with multiple medoids for large data	-	-	-	-	-	-	2	9	9	5	25
Incremental grid density-based clustering algorithm	6	2	3	2	2	1	2	1	5	0	24
An incremental clustering scheme for data de-duplication	-	-	3	1	2	2	5	5	2	2	23

Incremental clustering of mobile objects	2	2	1	1	2	2	2	1	6	2	21
Knowledge augmentation via incremental clustering: New technology for effective knowledge management	-	-	-	-	-	-	3	7	3	1	14
Evolve systems using incremental clustering approach	-	-	-	-	-	-	-	-	4	5	9

Source: http://www.scopus.com (accessed on 19th July 2018)

3.9 Source Statistics

From source statistics (Figure 14) for publications in incremental clustering, it is clear that maximum numbers of publications are from Lecture Notes in Computer Science (LNCS).

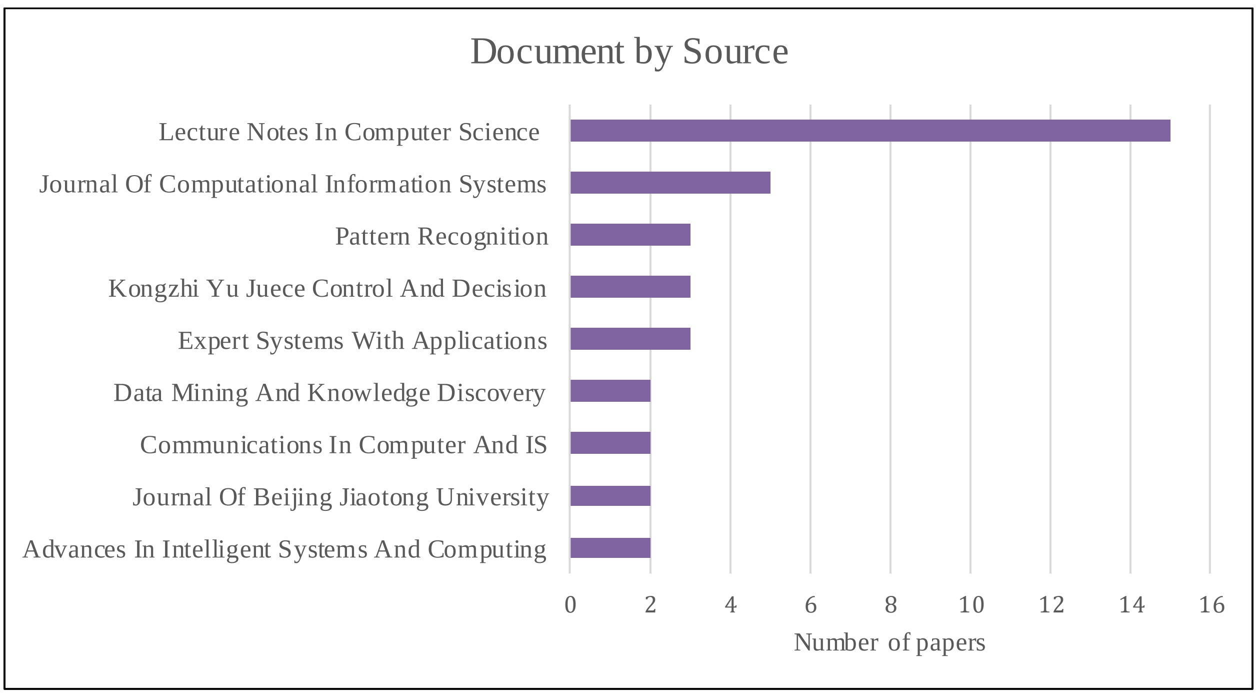


Figure 14: Source statistics for publications in incremental clustering

Source: http://www.scopus.com (accessed on 19th July 2018)

4. RESEARCH IMPLICATION OF THE STUDY

Incremental clustering research work is done worldwide, and it is continuously growing. This research work perhaps laid the firm groundwork that will prompt inventive, creative ability and enlighten the significance of incremental clustering process to bring in the change through improvement in their processes.

The bottom level keywords in table 6 of a bibliometric analysis of incremental clustering algorithms are artificial intelligence, database systems, learning system, probability, and neural network. This implies a major research gap which is consequently providing critical space to analysts around the globe to concentrate research in artificial intelligence, database systems, learning system, probability, and neural network. These domains have the more prominent scope and therefore needs to be investigated.

China and India are the world rising economies, and their research framework is growing at a quick pace. In the area of incremental clustering, only four review papers out of 173 exist. With the ever-increasing demand for handling data which is continuously updated, Indian researchers can explore this area which has been unnoticed in the past.

5. CONFINES OF THE PRESENT STUDY

This study explores the Scopus database using a combination of keywords for analysis purpose. Some vital publications and journal articles were not available in Scopus database during data analysis of this study hence they couldn't be incorporated in this study. This study also restricts research papers in English language only.

6. CONCLUSIONS

This bibliometric survey on incremental clustering algorithms revealed that maximum publications are of LNCS type. They are from conference and journals, affiliated with computer science. Chinese lead these publications followed by India then the US. Convergence optimality is another prominent keyword and less attentiveness towards correlation, betweenness and "friendly measures" keywords observed. After physics and astronomy - engineering is the contributing subject area and minimal contribution in terms of review papers.

The Incremental Clustering is the niche area of research related to Advanced Machine Learning. It is easy to carry out the incremental learning via incremental clustering related to huge data sets of varied domains, in a collaborative form. To achieve effectual incremental learning, algorithms such as CFBA, CBICA and alike; are required to be applied to various domains to build user-friendly systems, which are very useful for society at large. To achieve best quality clusters, it is necessary to build parameter-free incremental clustering algorithms. Such algorithms proved, maintained and ranked not only the clusters over various iterations dynamically but also the impactful attributes. In the preprocessing stage, the attributes are chosen by algorithms like PCA / ICA /FS / FE etc. can be easily proved at post-clustering stage using incremental clustering algorithms. The highly ranked cluster(s) or also termed as high priority clusters can be well utilized by authorities for effectual decision making, (for example) in case of load profiling, load shedding (planned and unplanned), floating solar plant related studies based on electricity smart-meter data analysis. Personalization in medical treatment can be implemented using such cluster ranking concepts. Incremental clustering is the most important and growing field of study and hence felt the need to undertake the detailed bibliometric study about this topic.

REFERENCES

Lu, Z. (Ed.). (2013). *Information retrieval methods for multidisciplinary applications*. IGI Global. https://www.igi-global.com/book/information-retrieval-methods-multidisciplinary-applications/72351

Swamy, B. K., & Kulkarni, P. A. (2006, July). Intelligent decision making based on pattern matching & mind-maps. In *Proceedings of the 10th WSEAS international conference on Computers* (pp. 492-497). World Scientific and Engineering Academy and Society (WSEAS). http://www.wseas.us/e-library/conferences/2006cscc/papers/534-947.pdf

Sarmiento, A. M., & Nagi, R. (1999). A review of integrated analysis of production-distribution systems. *IIE transactions*, *31*(11), 1061-1074. https://link.springer.com/article/10.1023/A:1007623508610

Gaikwad, S. M., Joshi, R. R., & Mulay, P. (2015). Cluster Mapping with the help of New Extended MCF Algorithm and MCF Algorithm to Recommend an Ice Cream to the Diabetic Patient. *METHODOLOGY*, *1*(6), 7. https://www.researchgate.net/publication/282876798_Cluster_Mapping_with_the_help_of_New_Extended_MCF_Algorithm_and_MCF_Algorithm_to_Recommend_an_Ice_Cream_to_the_Diabetic_Patient

Degadwala, S. D., Mahajan, A. D., & Vyas, D. J. Privacy Preservation using T-Closeness with Numerical Attributes. http://www.ijfrcsce.org/download/browse/Volume_3/November_17_Volume_3_Issue_11/1511848865_28-11-2017.pdf

Joshi, P. M., & Kulkarni, P. A. (2011). A novel approach for clustering based on pattern analysis. *International Journal of Computer Applications*, *975*, 8887.

https://www.ijcaonline.org/archives/volume25/number4/3023-4089

Mulay, P., Patel, K., & Gauchia, H. G. (2017). Distributed System Implementation Based on "Ants Feeding Birds" Algorithm: Electronics Transformation via Animals and Human. In *Detecting and Mitigating Robotic Cyber Security Risks* (pp. 51-85). IGI Global. https://www.igi-global.com/chapter/distributed-system-implementation-based-on-ants-feeding-birds-algorithm/180061

Mulay, P., Joshi, R. R., Anguria, A. K., Gonsalves, A., Deepankar, D., & Ghosh, D. (2017). Threshold Based Clustering Algorithm Analyzes Diabetic Mellitus. In *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications* (pp. 27-33). Springer, Singapore. https://link.springer.com/chapter/10.1007/978-981-10-3156-4_3

Mulay, P., & Kulkarni, P. A. (2013). Knowledge augmentation via incremental clustering: new technology for effective knowledge management. *International Journal of Business Information Systems*, *12*(1), 68-87. https://dl.acm.org/citation.cfm?id=2407446

Shinde, K., & Mulay, P. (2017, April). Cbica: Correlation based incremental clustering algorithm, a new approach. In *Convergence in Technology (I2CT), 2017 2nd International Conference for* (pp. 291-296). IEEE. https://ieeexplore.ieee.org/document/8226138

Lakshmanaprabu, S. K., Shankar, K., Gupta, D., Khanna, A., Rodrigues, J. J., Pinheiro, P. R., & de Albuquerque, V. H. C. (2018). Ranking analysis for online customer reviews of products using opinion mining with clustering. *Complexity*, *2018*. https://www.hindawi.com/journals/complexity/2018/3569351/

Lakshmanaprabu, S. K., Shankar, K., Khanna, A., Gupta, D., Rodrigues, J. J., Pinheiro, P. R., & De Albuquerque, V. H. C. (2018). Effective Features to Classify Big Data Using Social Internet of Things. *IEEE Access*, *6*, 24196-24204. https://ieeexplore.ieee.org/document/8349962

Kulkarni, P. A., & Mulay, P. (2013). Evolve systems using incremental clustering approach. *Evolving Systems*, *4*(2), 71-85. https://link.springer.com/article/10.1007/s12530-012-9068-z

Kulkarni, P. A., Dwivedi, S., & Haribhakta, Y. V. (2015). *U.S. Patent Application No. 14/676,680*.

Kadam, S., Bandyopadhyay, P. K., & Patil, Y. (2016). Mapping the field through bibliometric analysis of passenger centric railway transportation. *International Journal of Automation and Logistics*, *2*(4), 349-368. https://www.inderscienceonline.com/doi/abs/10.1504/IJAL.2016.080340

Kulkarni, H. (2017, September). Intelligent context based prediction using probabilistic intent-action ontology and tone matching algorithm. In *Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on* (pp. 656-662). IEEE. https://ieeexplore.ieee.org/document/8125916

Kulkarni, A., Tokekar, V., & Kulkarni, P. (2015). Discovering context of labeled text documents using context similarity coefficient. *Procedia computer science*, *49*, 118-127. https://www.sciencedirect.com/science/article/pii/S1877050915007449

Kumar, A., Ahuja, H., Singh, N. K., Gupta, D., Khanna, A., & Rodrigues, J. J. (2018). Supported matrix factorization using distributed representations for personalised recommendations on twitter. *Computers & Electrical Engineering*, *71*, 569-577. https://www.sciencedirect.com/science/article/pii/S0045790618312084

Joshi, R. R., & Mulay, P. (2018). Deep Incremental Statistical Closeness Factor Based Algorithm (DIS-CFBA) to assess Diabetes Mellitus. *BLOOD*, *115*, 210. https://www.irjet.net/archives/V5/i7/IRJET-V5I756.pdf

Johnson, T., & Singh, S. K. (2016). Quantitative Performance Analysis for the Family of Enhanced Strange Points Clustering Algorithms. *International Journal of Applied Engineering Research*, *11*(9), 6872-6880. https://pdfs.semanticscholar.org/d513/63dc76eea4bc276ea7fe54dc53aae294de7e.pdf

Gaikwad, S. M., Joshi, R. R., & Mulay, P. (2016). System dynamics modeling for analyzing recovery rate of diabetic patients by mapping sugar content in ice cream and sugar intake for the day. In *Proceedings of the Second International Conference on Computer and Communication Technologies* (pp. 743-749). Springer, New Delhi. https://link.springer.com/chapter/10.1007/978-81-322-2517-1_71

Archana Chaudhari, & Mulay, P. (2018). SCSI: Real-time Data Analysis with Casandra and Spark. *Big Data Processing Using Spark in Cloud* (Vol. 43). Studies in Big Data Springer. https://link.springer.com/chapter/10.1007/978-981-13-0550-4_11